



Konventionaliserade fraser i en akademisk ordlista

Forskning kring Academic Formulas List (AFL)

Ellis, Simpson-Vlach & Maynard 2008

Simpson-Vlach & Ellis 2010



Varför en AFL?

- “Research and practice in second language education demonstrates that academic study puts substantial demands on students because the relevant language necessary for proficiency in academic contexts is quite different from that required basic interpersonal communicative skills” (Ellis et al. 2008:379).
- “Recent research in corpus linguistics [...] has established that highly frequent formulaic expressions are not only salient but also functionally significant: Cognitive science demonstrates that knowledge of these formulas is crucial for fluent processing” (Ellis et al. 2008:379).



Forskning med målet att skapa en "Academic Formula List" (AFL)

- AFL tar upp konventionaliserade fraser (formulaic sequences) som
 - Har identifierats som frekventa återkommande mönster i tal- och skriftspråskorpusar
 - Som är mer frekventa i akademisk än i icke-akademisk diskurs
 - Används på många olika akademiska ämnesområden

Ellis, Simpson-Vlach & Maynard 2008:

- Excerperingen av 'pedagogiskt användbara' konventionaliserade uttryck för akademisk tal och skrivande med hjälp av korpuslingvistiska metoder
- Ett bedömningstest med lärare i engelska som andraspråk (ESL) och "english for academic purposes" (EAP)
 - Värdering av de excerperade uttryckens pedagogiska värde
- Tre experimentella undersökningar "which show that different aspects of formulaicity affect the accuracy and fluency of processing of these formulas in native speakers and in advanced L2 learners of English" (Ellis et al. 2008:375)

Korpusstudien

- Akademiska tal- och skriftspråskorpusar bestående av sammanlagt 2,1 mio ord
 - Akademisk tal: MICASE (1,7 mio ord) (Simpson et al.2002) och utvalda filer från BNC (431 000 ord)
 - Akademisk skrift: Hyland's (2004) research article corpus och utvalda filer från BNC (931 000 ord)
 - Två refrens korpusar med icke-akademiskt tal och skrift:
 - » Switchboard (2006) - informella telefonsamtal (2,9 mio ord)
 - » FLOB och Frown (1991) - skrift, brittisk och amerikansk engelska, 15 olika genre (1,9 mio ord)

Korpusstudien

- 3-grams, 4-grams och 5-grams med en frekvens av minst 10 per 1 miljon ord
- Collocate (Barlow 2004) användes för att mäta både frasernas frekvens och MI-värden (Mutual Information)
 - Högre MI-värde betyder att det finns en starkare association mellan orden som ingår i frasen, ett lägre värde betyder att samförekomsten är med större sannolikhet en slump (skala, inget signifikanstest)
 - Varken frekvens eller MI för sig är lämpliga mått för att identifiera fraser som är "highly noteworthy for teachers" (Simpson-Vach & Ellis 2010:493)
 - » Tabell 2 och 3 på stödpappret



Korpusstudien

- Sammanlagt ca 14 000 fraser som förekommer minst 10 gånger per 1 miljon ord i någon av korpusarna
- De fraser som är statistiskt mer frekventa ($p < 0.01$) i de akademiska korpusarna identifierades med hjälp av log-likelihood (statistisk jämförelse av frasernas relativa frekvenser mellan de akademiska och icke-akademiska korpusarna).
 - Akademisk tal vs icke akademisk tal
 - » Drygt 2000 fraseologiska enheter
 - Akademisk skrivande vs icke-akademisk skrivande
 - » Nästan 2000 fraseologiska enheter

Lärarnas bedömning

- Subsample av 108 fraser (54 tal, 54 skrift)
- Är de excerperade fraser relevanta ur ett pedagogiskt perspektiv, dvs. Bl.a. för undervisning i EAP?
 - lärare i ESL och EAP har fått göra en av följande bedömningar:
 - » (1) om de ansåg ordkombinationerna ifråga utgöra fasta fraser, konventionaliserade uttryck eller 'chunks'
 - » (2) om de ansåg fraserna ha en betydelse och funktion som en enhet (som inte motsvarar de enskilda konstituenttryck ternas betydelse och funktion)
 - » (3) om de ansåg fraserna vara värda att läras ut som idiomatiska uttryck
 - » Skala 1 (instämmer inte) till 5 (instämmer helt)

Kan bedömningar förutsägas utifrån frasernas frekvens eller MI

- Korrelationsanalys visar att båda värden spelar roll för bedömningen – men MI är en mer avgörande faktor
- Framgångsrik triangulering av korpuslingvistiska värden och lärarnas bedömningar
 - Frekvens
 - MI
 - Värden från ‘instructor insights’- analys
 - » Formula Teaching Worth (FTW-score) som kan förutsäga hur en lärare hade bedömt frasen i fråga

Psykolingvistiska experiment – processhastighet för L1- och L2-talare

- a) Leshastighets – och grammatikalitetsbedömningstest
 - Hälften av fraserna “äka”, hälften nonsense
- b) Högläsning av fraser som kommer upp på en skärm en i taget – Voice Onset & Articulation Time
 - mätning av pausen mellan att frasen kommer upp på skärmen och början på informantens talade respons
- c) ‘Priming’ av sista ordet i en fras
 - mäter i vilken mån informanten får snabbare tillgång till uttalet av det sista ordet genom att se början på frasen

Resultaten av alla tre experiment relateras till ordlängd, frekvens och MI



Slutsats (relevant med tanke på AFL-användare)

“...native speakers and advanced ESL learners have become sensitive from their usage histories to these [formulaic] expressions so that they process them preferentially. But native speakers and learners are sensitive to different determinants of fluency – learners to [...] frequency, fluent natives to MI” (Ellis et al. 2008:389).



Resultat: The AFL

- Tre 'sublists'
 - 'Core AFL' (tal och skrift)
 - Spoken AFL
 - Written AFL
- Sorterade efter FTW-score



Funktionella kategorier

Klassificering i pedagogisk syfte (Simpson-Vlach & Ellis 2010:497ff):

- A. Referential expressions
- B. Stance expressions (epistemiska uttryck?)
- C. Diskourse organizing functions

→ Ett antal Olika subkategorier

(se tabell 4 på stödpappret)



Referenser

Ellis, Nick. C., Simpson-Vlach, Rita & Maynard, Carson (2008).
Formulaic Language in Native and Second Language Speakers:
Psycholinguistics, Corpuslinguistics and TESOL. *TESOL Quarterly*
42/3: 375–396.

Simpson-Vlach and Ellis (2010). An Academic Formulas List: New
Methods in Phraseology Research. *Applied Linguistics* 31/4: 487-512.